JOURNAL OF VIROLOGY, Sept. 1979, p. 752-760
0022-538X/79/09-0752/09$02.00/0

Vol. 31, No. 3

Heteroduplex Analysis of the Sequence Relationships Between the Genomes of Kirsten and Harvey Sarcoma Viruses, Their Respective Parental Murine Leukemia Viruses, and the Rat Endogenous 30S RNA

YUEH-HSIU CHIEN,[1] MICHAEL LAI,[2] THOMAS Y. SHIH,[3] INDER M. VERMA,[4] EDWARD M. SCOLNICK,[3] PRADIP ROY-BURMAN,[2] AND NORMAN DAVIDSON[1]*

Department of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125[1]; Departments of Microbiology and Pathology, University of Southern California School of Medicine, Los Angeles, California 90033[2]; Laboratory of Tumor Virus Genetics, National Cancer Institute, Bethesda, Maryland 20205[3]; and The Salk Institute, San Diego, California 92112[4]

Received for publication 28 February 1979

The sequence relations between Kirsten murine sarcoma virus (Ki-SV), Harvey murine sarcoma virus (Ha-SV), and a rat endogenous 30S RNA were studied by electron microscope heteroduplex analysis. The sequence relationships between the sarcoma viruses and their respective parental murine leukemia viruses (Kirsten and Moloney murine leukemia viruses), as well as between the two murine leukemia viruses, were also studied. The only observed nonhomology feature of the Kirsten murine leukemia virus/Moloney murine leukemia virus heteroduplexes was a substitution loop with two arms of equal length extending from 1.80 ± 0.18 kilobases (kb) to 2.65 ± 0.27 kb from the 3′ end of the RNA. It is believed that this feature lies in the *env* gene region of the viral genomes. The Ha-SV and Moloney murine leukemia virus genomes (respective lengths, 6.0 and 9.0 kb) were homologous in a 1.0 ± 0.05-kb region at the 3′ end and possibly over a 200-nucleotide region at the 5′ ends; otherwise, they were nonhomologous. Ha-SV and Ki-SV (length, 7.5 kb) were homologous in the first 4.36 ± 0.37-kb region from the 3′ end and in a 0.70 ± 0.15-kb region at the 5′ end. In between, there was a nonhomology region, possibly containing a short (0.23-kb) region of partial or total homology. The heteroduplex analysis between rat endogenous 30S RNA and Ki-SV shows that there are mixed regions of sequence homology and nonhomology at both the 5′ and 3′ ends. However, there is a large (4-kb) region of homology between Ki-SV and the rat 30S RNA in the center of the genomes, with only a small nonhomology hairpin feature. These studies help to define the regions of homology between the Ha-SV and Ki-SV genomes with each other and with the rat endogenous 30S RNA. These regions may be related to the sarcoma genicity of the viruses. In particular, the 0.7-kb region of homology of Ha-SV with Ki-SV at the 5′ ends may be related to the formation of a 21,000-dalton phosphoprotein in cells transformed by either virus.

Kirsten murine sarcoma virus (Ki-SV) and Harvey murine sarcoma virus (Ha-SV) were isolated from tumors induced by inoculation of Kirsten murine leukemia virus (Ki-MLV) and Moloney murine leukemia virus (Mo-MLV) into rats (8, 9). The sarcoma virus genomes are defective in replication, but they can induce solid tumors in animals and transform fibroblasts in tissue culture (1, 11, 15). A number of studies have demonstrated that the genomes of these sarcoma viruses contain some sequences homologous to the respective parental leukemia viruses and sequences derived from rat genetic information (2, 5, 13, 15, 16). Recent studies have shown that this rat genetic information is expressed in a variety of rat cells as an approximately 30S RNA (14, 17, 20).

The location of the mouse and rat genetic information in Ki-SV and Ha-SV has been studied by oligonucleotide fingerprinting (18, 19). For Ki-SV, the residual Ki-MLV-related sequences were located in the terminal 1,000 bases at the 3′ end of the virus and in a short stretch at the very 5′ end. Between the parental mouse leukemia virus genetic information of Ki-SV lies a large insert of approximately 6.5 kilobases (kb), which was shown to contain rat-derived sequences throughout its entire length. In Ha-SV,

a large portion of the rat-related sequences was shown to be homologous to the rat insert in Ki-SV. In this paper, the sequence relationships among Ha-SV, Ki-SV, and the rat endogenous RNA were studied by electron microscopic heteroduplex analysis. The sequence relationships between the sarcoma viruses and their respective parental murine leukemia viruses, as well as between the two murine leukemia viruses, were also studied. More detailed maps were determined, which more precisely localize the regions of homology and nonhomology among the viruses.

MATERIALS AND METHODS

Ha-SV RNA was prepared from Ha-SV/Mo-MLV pseudotype virions released from a Fisher rat embryo cell (FRE clone 2) transformed by Ha-SV (18). This culture produced a stable 15:1 excess of Ha-SV over Mo-MLV. Viruses were harvested at 2-h intervals and purified by banding on a sucrose density gradient. The pseudotype virions were disrupted in TNE buffer (0.01 M Tris-hydrochloride [pH 7.4], 0.1 M NaCl, and 1 mM EDTA) containing 1% sodium dodecyl sulfate and 0.2% (vol/vol) diethylpyrocarbonate. 60 to 70S viral RNA was isolated by sedimentation through a 15 to 30% sucrose gradient in TNE buffer at 17,000 rpm in an SW27 rotor for 17 h at 20°C. Polyadenylated [poly(A)$^+$] RNA was prepared from heat-dissociated 60 to 70S RNA as previously described (6).

The HTP cells containing high levels of rat endogenous 30S RNA were originally derived from a chemically induced uterine carcinoma of a Sprague-Dawley rat. The endogenous 30S RNA was rescued by superinfection with woolly leukemia virus (17). This culture produces equal amounts of 30S RNA and woolly leukemia virus. The pseudotype virions were harvested at 2-h intervals. The endogenous rat 30S RNA in the pseudotype virions was isolated by the same procedure described for Ha-SV RNA.

Ki-MLV was propagated on SC-1 mouse cells (19). The 70S RNA was isolated by published procedures (10). Ki-SV was harvested from a transformed NRK line (no. 58967) which produces a 50-fold excess of Ki-SV over its helper (11). The passage history of this line has been described previously (13).

The sample of Mo-MLV complementary DNA (cDNA) that has been described previously (6) was used in the present study. The lengths of these molecules, as determined by electron microscopy, were distributed approximately uniformly between 6 and 10 kb. cDNA of Ki-SV was synthesized by published procedures (10). The purified cDNA was separated by alkaline sucrose gradient sedimentation, and only the DNA equal to or larger in length than simian virus 40 (SV40) DNA was used for this study. The fraction of molecules in different length intervals in the Ki-SV cDNA preparation were estimated by electron microscopic observation to be as follows: 10%, 4 to 5 kb; 13%, 5 to 6 kb; 30%, 6 to 7 kb; 2%, 7 to 8 kb; 20%, 8 to 9 kb; 7%, 9 to 10 kb. Some of the cDNA in the 8- to 9-kb interval may be Ki-SV (RNA genome length, 7.5 ± 0.3 kb by gel electrophoresis); the rest is the helper Ki-MLV.

RNA:cDNA hybridization was carried out as follows. A mixture of 1.5-μg/ml 30 to 35S RNA and 1-μg/ml cDNA in 80% formamide–0.4 M NaCl–0.01 M PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)]–0.001 M EDTA, pH 6.3, was incubated at 50°C for 0.5 h. Samples were incubated with SV40 bromodeoxyuridine (BUdR) for mapping the poly(A) ends of the RNA (4). From electron microscopy, we estimated that the lengths of the poly(BUdR) tails were less than 0.1 kb.

Heteroduplexes were spread from 60% formamide for electron microscopy as described previously (6).

RESULTS

Molecular length and composition of RNA preparations. Electrophoresis in 1% agarose–5 mM CH_3HgOH gels (3) showed that the Ha-SV RNA preparation contained two discrete bands with molecular lengths of 6.0 ± 0.3 and 9.0 ± 0.5 kb, corresponding to the sarcoma and the helper Mo-MLV components, respectively (Fig. 1, lane C). The Ki-SV RNA preparation contained bands with lengths of 7.5 ± 0.3 and 9.0 ± 0.5 kb, corresponding to the sarcoma and the helper Ki-MLV components, respectively (Fig. 1, lane B). The rat endogenous viral 30S RNA preparation had two bands with molecular lengths of 8.0 ± 0.3 and 9.0 ± 0.5 kb, corresponding to the rat endogenous 30S RNA and its helper woolly leukemia virus, respectively (Fig. 1, lane G).

Ki-MLV RNA/Mo-MLV cDNA heteroduplexes. Heteroduplexes were prepared as described above and spread from 60% formamide. A typical heteroduplex molecule is shown in Fig. 2a. A schematic representation of the heteroduplex structure, with a summary of the data, is given in Fig. 3. The only observed nonhomology feature was a substitution loop with two arms with measured lengths of 0.75 ± 0.30 kb (n = 54) and 0.62 ± 0.25 kb (n = 52). These two murine leukemia viruses appeared to be homologous at the heteroduplex criterion over the remainder of their genomes. Note that single-strand lengths were relative to a φX174 single-strand DNA standard. Previous studies (6) have shown that RNA single strands are more condensed and shorter than DNA single strands by about 18% under the spreading conditions used here. If we assume that in the present case the arm with a measured length of 0.62 kb was RNA, its corrected length was 0.75 kb, and the two arms of the nonhomology loop were approximately equal in length. With this interpretation, which was adopted in the presentation of the data in Fig. 3, the region of sequence nonhomology between the two genomes extended from 1.80 ± 0.18 kb (n = 46) to 2.65 ± 0.27 kb (n = 44) from the 3′ end of the RNA.

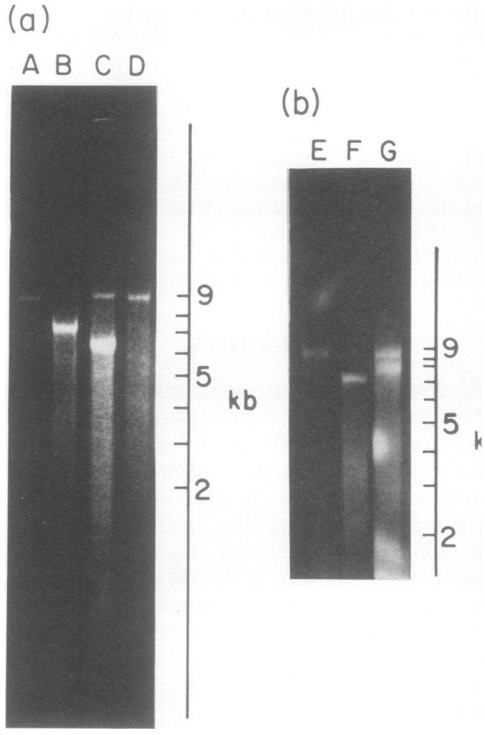

FIG. 1. *(a) Electrophoretic analysis of viral RNAs on CH_3HgOH–1% agarose gels: (A) Mo-MLV 35S RNA; (B) Ki-SV; (C) Ha-SV; (D) Ki-MLV. The molecular weight standards are Mo-MLV, Escherichia coli 23S, and 16S rRNA's (data not shown) (molecular lengths of 9.0, 3.1, and 1.7 kb, respectively). (b) Analysis of rat endogenous 30S RNA on CH_3HgOH–1% agarose gels: (E) Mo-MLV 35S RNA; (F) Ki-SV; (G) rat endogenous 30S RNA. The stained areas below 5 kb in (G) are not due to RNA, but are artifacts due to nonuniform staining of the gel. The molecular weight standards are Mo-MLV, E. coli 23S, and 16S rRNA's (data not shown) (molecular lengths of 9.0, 3.1, and 1.7 kb, respectively).*

One important reservation about the interpretation of the heteroduplex structure shown in Fig. 2a should be noted. Gel electrophoresis showed that the Ki-MLV RNA preparation contained some full-length RNA, but a significant mass fraction of broken molecules. Furthermore, not all of the cDNA molecules were of full length. Therefore, the number of heteroduplex molecules with both strands intact near the 5′ end of the RNA was less than that of heteroduplexes with the 3′-proximal strands intact. Many heteroduplexes which appeared in the electron microscope to be duplex from the substitution loop to the 5′ end of the genome were seen. However, single-strand RNA or DNA versus RNA:DNA hybrid discrimination in these electron micrographs was not always unambiguous. It is therefore possible that some small nonhomology features near the 5′ end of these heteroduplexes may have escaped detection.

Ha-SV RNA/Mo-MLV cDNA heteroduplexes. A typical heteroduplex structure is shown in Fig. 2b. A region of homology with a length of 1.0 ± 0.05 kb (n = 80) at the 3′ end of the RNA and two long single-strand arms could be seen. In three such molecules, the two long single-strand arms came together at their distal ends in what appeared to be a duplex region with a length of about 200 nucleotide pairs or less, and the two single-strand arms were of the correct length to be assigned to full-length Ha-SV RNA and Mo-MLV cDNA. The proposed heteroduplex structure is shown in Fig. 3b. The number of structures observed was not sufficient to prove the interpretation that Ha-SV and M-MLV share a very short homology region at the 5′ end. Therefore, this proposed homology region is shown as dotted lines.

For reasons explained below, we believe that the Ki-SV/Ki-MLV heteroduplex should have a structure similar to that of Ha-SV/Mo-MLV. Observations of heteroduplexes with large nonhomology loops and small duplex regions, such as that shown in Fig. 2b, require full-length molecules; however, because a satisfactory, unbroken Ki-MLV RNA preparation was not available, we were unable to obtain decisive results.

Ha-SV RNA/Ki-SV cDNA. When heteroduplex preparations were spread from 60% formamide, two types of molecules were seen with approximately equal frequency. Micrographs are shown in Fig. 2c and d, and schematic representations of the structures are shown in Fig. 3cA and B. In both cases, there was a duplex region with a length of 4.36 ± 0.37 kb (n = 46) at the 3′ end and a duplex segment with a length of 0.70 ± 0.15 kb (n = 46) at the 5% end of the RNA. Between these two duplex segments there was a region of nonhomology. In one case, it appeared as a completely open substitution loop with arms with lengths of 2.55 ± 0.30 kb (n = 26) and 1.07 ± 0.28 kb (n = 26) (Fig. 3cA). In about 80% of the structures of this type, there was a hairpin structure in the middle of the shorter arm with a measured length of 0.34 kb pairs. We observed an approximately equal number of molecules with the structure shown in Fig. 3cB. In this case, the nonhomology region consisted of two substitution loops with the dimensions shown in the figure which were separated by a duplex segment with a length of 0.23 ± 0.04 kb (n = 20). The longer arms of the substitution loops shown in Fig. 3c were assigned to Ki-SV, to make the heteroduplex data con-

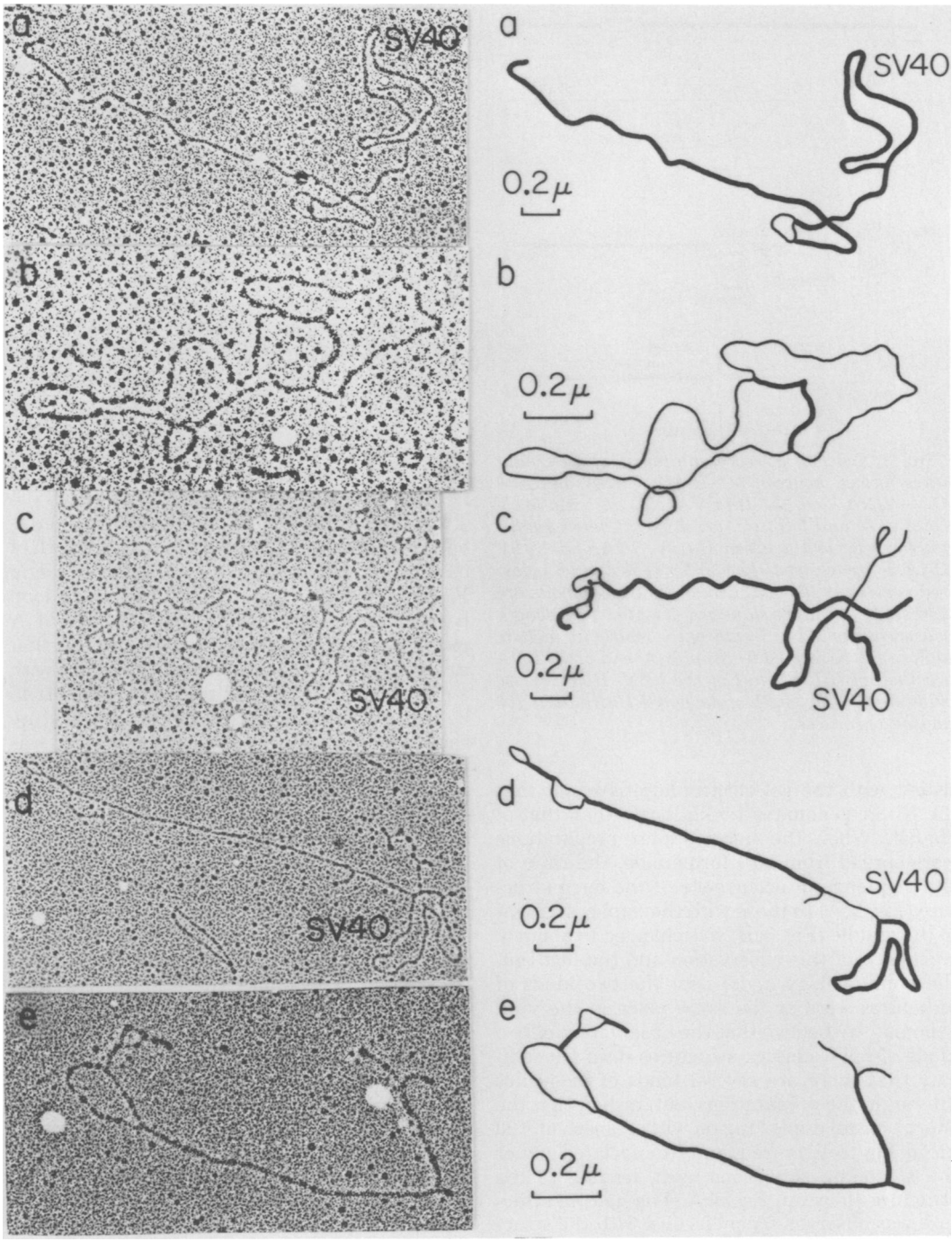

FIG. 2. *Electron micrographs of (a) linear heteroduplex molecule formed between Mo-MLV cDNA and Ki-MLV RNA (a circular SV40 BUdR molecule is attached at the 3′ end of the viral RNA in the heteroduplex); (b) circular heteroduplex formed between Mo-MLV cDNA and Ha-SV RNA; (c) linear heteroduplex formed between Ki-SV cDNA and Ha-SV RNA, with a broken SV40 BUdR molecule attached at the 3′ end of the RNA (short RNA molecules with intact 3′ ends are also attached to the SV40 BUdR molecule); (d) linear heteroduplex molecule formed between Ki-SV cDNA and rat endogenous 30S RNA, with a circular SV40 BUdR molecule attached at the 3′ end of the RNA; (e) linear heteroduplex molecule formed between Ki-SV cDNA and rat endogenous 30S RNA. Interpretative tracings of the heteroduplexes are shown next to the micrographs. Heavy lines indicate duplex regions; thin lines represent single-strand regions.*

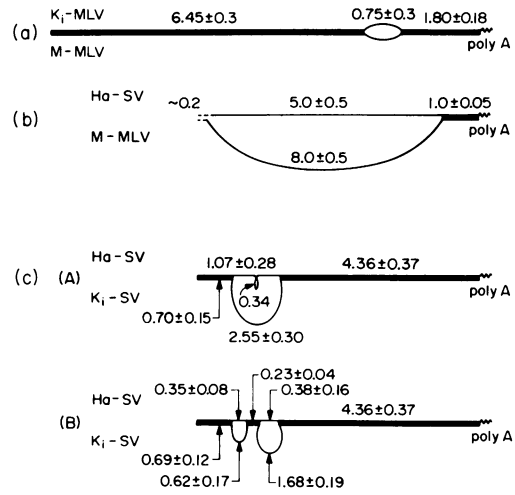

FIG. 3. *Schematic representations of heteroduplexes formed between (a) Ki-MLV RNA and Mo-MLV cDNA and (b) Ha-SV RNA and Mo-MLV cDNA. (cA) and (cB) are two different heteroduplex structures formed between Ha-SV RNA and Ki-SV cDNA. For each type, the number of molecules measured is given in the text. Lengths of the segments are expressed in kb ± the standard deviation. The length measurements at the 3' ends of the molecules include duplex regions formed by the poly(A) end of the RNA and the poly(BUdR) tail of the label. However, we estimate that the length of the poly(BUdR) tail is 100 nucleotides or less.*

sistent with the gel electrophoretic result that the Ki-SV genome is 1.5 kb longer than that of Ha-SV. When the heteroduplex preparations were spread from 80% formamide, the ratio of the frequency of occurrence of the open structure (Fig. 3cA) to those with the duplex segment in the middle (Fig. 3cB) was changed to about 3:1. In view of this observation and the fact that the nonhomology regions for the two kinds of structures were at the same place in the viral genomes, we believe that the observation of two kinds of heteroduplex structures does not indicate that there are several kinds of molecules present in the preparations, but, rather, that the short, central duplex region with a length of 0.23 kb in Fig. 3cB is frequently dissociated under the spreading conditions used, leading to the structure shown in Fig. 3cA. The hairpin structure was observed for molecules with the structure shown in Fig. 3cA but not for that shown in Fig. 3cB. The possibility of forming such a hairpin would decrease the probability that a structure of the type shown in Fig. 3cA would be converted to one like that shown in Fig. 3cB.

Rat endogenous 30S RNA/Ki-SV cDNA. Several different kinds of heteroduplex structures were observed. Electron micrographs of two of them are shown in Fig. 2d and e. Representations of the several heteroduplex structures are shown in Fig. 4. Because of the complexity of the results, the discussion is divided into separate sections for different regions of the genome.

(i) 5' end. Four different kinds of heteroduplex structures were observed at the end corresponding to the 5' end of the RNA; these are labeled as Fig. 4a, b, c, and d. In all cases there was a relatively large substitution loop with arms $a'_1 = 1.80 \pm 0.26$ kb ($n = 54$) and $a_1 = 0.62 \pm 0.2$ kb ($n = 54$). The lengths of these features were the same regardless of whether molecules were of the type shown in Fig. 4a, b, c, or d. Proceeding toward the 5' end, in Fig. 4a, b, and c there was a short duplex d_3 with a length of 0.23 ± 0.05 kb ($n = 22$). In the structure shown in Fig. 4a, proceeding from the duplex d_3 toward the 5' end, there was a substitution loop with arms of lengths $b_1 = 0.51 \pm 0.18$ kb ($n = 6$) and $b'_1 = 0.46 \pm 0.19$ kb ($n = 6$). The 5' end of the heteroduplex appeared as a duplex d_4 with a length of 0.31 ± 0.19 kb ($n = 6$). In the structure shown in Fig. 4b, the duplex region d_3 was joined to a loop b with a total length of 1.54 ± 0.07 kb ($n = 5$). A plausible interpretation of this structure is that it was similar to that shown in Fig. 4a, but with a duplex stem at the end that was too short to be seen by electron microscopy. In the structure shown in Fig. 4c, the duplex segment d_3 was terminated by a single-strand fork. The fourth class of structure was a segment d_5 with a length of 0.68 ± 0.25 kb ($n = 54$) and no observed sequence nonhomology features. The presence of these four classes of molecules may indicate that the rat endogenous 30S RNA is heterogeneous in sequence, but it is also possible that the different structures shown in Fig. 4a, b, c, and d all arose from heteroduplexes with a single RNA sequence. For example, there may have been partial sequence mismatch in the d_4 duplex region in the structure shown in Fig. 4a. It could be opened up under the spreading condition and give rise to the b loop in the structure shown in Fig. 4b, with a duplex region with a length of fewer than 50 nucleotides at the very 5' end. If the b loop opened completely or if the RNA or the cDNA or both were not long enough, the structure shown in Fig. 4c was observed. When the RNA or the cDNA or both were not long enough to cover the region d_3, the structure shown in Fig. 4d could be observed. It is also possible that the structures shown in Fig. 4a, b, and c were all heteroduplexes of the Ki-SV cDNA with a single 30S RNA sequence, but the structure shown in Fig. 4d was another class of molecule, where there was a sequence homology region with a length of 0.68 kb at the very 5' end

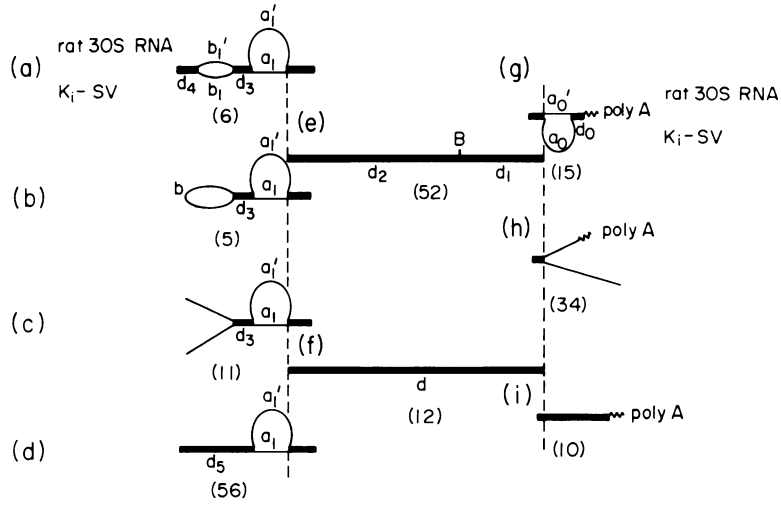

FIG. 4. *Schematic representation of the structures observed in rat endogenous 30S RNA/Ki-SV heteroduplexes. (a), (b), (c), and (d) are structures observed at the 5′ end. (e) and (f) are structures observed in the middle of the molecules. (g), (h), and (i) are structures observed at the 3′ end. The number of molecules observed with a particular structure is given within parentheses below that structure. (i) was scored only when there was a circular SV40 BUdR molecule attached. The length measurements were as follows:* d_0 = 0.36 ± 0.25 kb (n = 15); d = 4.03 ± 0.35 kb (n = 12); d_1 = 1.28 ± 0.07 kb (n = 15); d_2 = 2.55 ± 0.19 kb (n = 47); d_3 = 0.25 ± 0.05 kb (n = 22); d_4 = 0.31 ± 0.19 kb (n = 6); d_5 = 0.68 ± 0.25 kb (n = 56); a_0 = 1.35 ± 0.50 kb (n = 15); a'_0 = 0.50 ± 0.43 kb (n = 15); a_1 = 0.62 ± 0.20 kb (n = 54); a'_1 = 1.80 ± 0.26 kb (n = 59); b_1 = 0.51 ± 0.18 kb (n = 6); b'_1 = 0.46 ± 0.19 kb (n = 6); b = 1.54 ± 0.07 kb (n = 5). *The hairpin B has a single-strand length of 0.2 ± 0.04 kb (n = 52). Note that the* d_0 *segment includes the duplex region formed by the poly(A) end of the RNA and the poly(BUdR) tail of the label.*

of the heteroduplex molecules.

(ii) Middle of the heteroduplex. On the 3′ side of the a_1/a'_1 substitution loop, there was a duplex region d_2 with a length of 2.55 ± 0.19 kb (n = 47). In 80% of the molecules observed, d_2 was followed by a hairpin B with an estimated total single-strand length of 0.20 ± 0.04 kb (n = 52) and then by another duplex region d_1 with a length of 1.28 ± 0.07 kb (n = 15) (Fig. 4e). In about 20% of the cases, the hairpin feature B was not observed. Instead, a continuous duplex region d with a total length of 4.03 ± 0.35 kb (n = 12) was observed (Fig. 4f). The lengths of d_1 and d were only measured for molecules with single-strand forks or a substitution loop at the 3′ end of the duplexes. As discussed previously, the observation of two classes of heteroduplex molecules may reflect the heterogeneity of the RNA population. The fact that the length of d is equal to the sum of the lengths d_1, d_2, and the single-strand length of B suggests that the two kinds of molecules are not due to heterogeneity in sequence, but because the hairpin B is rather unstable and not formed when one of the two strands of the heteroduplex is broken around this region.

(iii) 3′ End. The several structures observed at the 3′ end of the RNA are shown in Fig. 4g, h, and i. In the structures shown in Fig. 4g, the substitution loop had arms with lengths a_0 = 1.35 ± 0.50 kb (n = 15) and a'_0 = 0.50 ± 0.43 kb (n = 15), followed by a duplex segment d_0 with a length of 0.35 ± 0.25 kb (n = 15). This length included the poly(A) at the 3′ end of the viral RNA and the BUdR tails on the label. In the structure shown in Fig. 4h, there were two single-strand arms, but no duplex segments at the 3′ end; in the structure shown in Fig. 4i, there was only a duplex region. Again, this may reflect heterogeneity in the RNA population, but the structures shown in Fig. 4h and i could also be variations of the structure shown in Fig. 4g. Since the d_0 region included the duplex regions formed by the poly(A) end of the RNA and the poly(BUdR) tail of the label, the sequence homology region between cDNA and RNA in d_0 may be very short. It could be opened up to give the structure shown in Fig. 4h. Heteroduplex molecules with broken cDNA at the region corresponding to the 3′ end of the RNA would give rise to the structure shown in Fig. 4i. Fingerprinting data (19) suggest that Ki-SV and its parental virus Ki-MLV are homologous over a region of about 1 kb at the 3′ end, but not homologous with the rat endogenous 30S RNA in this region. Therefore, it was necessary to assign the segment a'_0 to the rat endogenous 30S RNA. To be consistent with the molecular

weights determined by gel electrophoresis, we then assigned a_1 and b_1 to Ki-SV and a'_1 and b'_1 to rat endogenous 30S RNA.

Finally, it should be noted that hybridization studies have shown that there is no detectable sequence homology among the helper RNA of woolly leukemia virus, the rat 30S preparation, and Ki-SV (19). Therefore, none of the complex structures observed could be attributed to the woolly leukemia virus RNA component.

DISCUSSION

It is important to reemphasize the scope and limitations of heteroduplex analysis. Heteroduplexes are normally spread under conditions that are only moderately denaturing. Regions which appear as nonhomologous (insertion or deletion loops or substitution loops) are regions where there is little or no sequence homology between the two genomes. However, regions which appear as a duplex may be mismatched by an amount of 10 to 25% (6); that is, they are certainly related in sequence, but not necessarily identical. Examples are known where by oligonucleotide fingerprinting patterns two type C RNA viral genomes are greatly different, but by heteroduplex analysis they show many regions of homology (6). The advantage of heteroduplex analysis is that it maps the regions of partial or complete sequence homology and those of nonhomology rather precisely along the genome.

It should also be noted that under the spreading conditions used here, single-stranded and double-stranded nucleic acids cannot be distinguished unambiguously all of the time. Consequently, certain structures cannot be fully interpreted with confidence.

Our present results may be summarized as shown in Fig. 5 and as follows.

(i) The major sequence nonhomology region between Ki-MLV and Mo-MLV extends from 1.8 to 2.55 kb from the 3′ end (Fig. 3a). On the basis of previous mapping studies of murine viruses, this region can be assigned to the *env* gene (6, 7, 12).

It has further been suggested in studies of the heteroduplexes of mink cell focus-inducing viruses with their parental murine leukemia viruses (6) that the 3′ terminus of the 1.9-kb gene for the 70,000-molecular-weight nonglycosylated polypeptide precursor *env* gene product lies at 1.4 kb or closer to the 3′ end of the RNA. Therefore, the nonhomology loop corresponds to an internal part of the polypeptide chain or a region including the amino terminus. It is interesting, and possibly not merely coincidental, that the position and length of the nonhomology region of Ki-MLV and Mo-MLV are approximately the same as those for the several mink cell focus-inducing viruses with their parental Mo-MLV's.

(ii) Our measurements show that Ha-SV and its parental virus, Mo-MLV, are homologous in a region with a length of 1.0 kb at the 3′ end (Fig. 3b). A previous oligonucleotide fingerprinting study has indicated that the same is true for Ki-SV and Ki-MLV (19). We reported above a few observations which suggest but do not prove that there may be a short region of homology between Ha-SV and Mo-MLV at the 5′ ends.

(iii) Figure 3c shows that Ki-SV and Ha-SV are related in sequence everywhere except for a nonhomology region starting 0.7 kb from the 5′ end and extending 2.55 kb on the Ki-SV genome and 1.07 kb on the Ha-SV genome. Within this region of sequence nonhomology, there is a 0.23-kb region of homology which also corresponds to a hairpin structure in the Ha-SV genome (Fig. 3c).

The genetic composition of these two sarcoma viruses confirms earlier studies, showing that neither virus codes for any proteins of the paren-

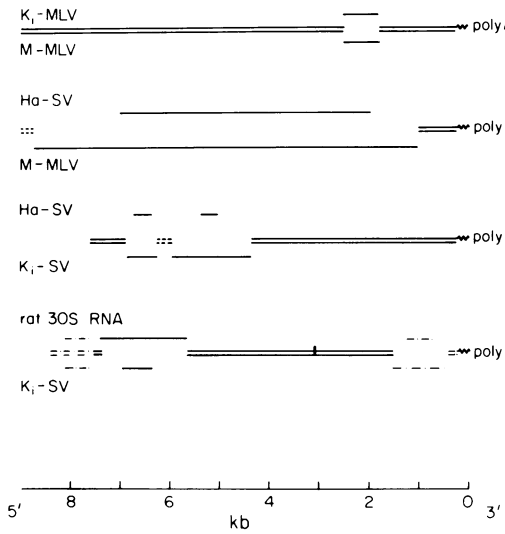

FIG. 5. *Summary of heteroduplex structures for Ki-MLV/Mo-MLV, Ha-SV/Mo-MLV, Ha-SV/Ki-SV, and rat endogenous 30S RNA/Ki-SV. Double lines indicate regions of sequence homology; single lines above or below the duplex double lines indicate sequences present in one and not in the other virus (sequence nonhomology regions). The broken lines in the Ha-SV/Mo-MLV and Ha-SV/Ki-SV heteroduplexes indicate regions which are partial sequence mismatches or an unstable short duplex region, as discussed in the text. All of the structures seen in rat endogenous 30S RNA/Ki-SV heteroduplexes are combined. —·—· lines indicate regions of multiple structure features, as discussed in the text.*

tal mouse type C viruses. No sequences are present in either Ki-SV or Ha-SV which code for *gag*, *pol*, or *env* gene products of either Ki-MLV or Mo-MLV. The genetic structure of both of these viruses suggests that only the very 5′ end and 3′ 1 kb of helper-independent type C viruses are necessary to provide the signals needed to propagate defective forms of type C retroviruses as pseudotypes. The structure suggests the possibility that perhaps any gene of the proper length could be inserted between these 5′ and 3′ ends and then become a transmissible retrovirus-like genome.

Furthermore, the structure suggests that any proteins coded for by Ki-SV or Ha-SV must derive from the insert between the 5′ and 3′ mouse type C sequences.

One important result of the heteroduplex studies is that there is a 0.7-kb region of sequence homology between Ki-SV and Ha-SV at the 5′ ends of the genomes. This region was not detected in a previous T1 oligonucleotide fingerprinting study (18). It is larger by at least 0.5 kb than the Mo-MLV-derived regions of the genomes at the extreme 5′ end.

Recent studies have shown that by using the antisera from animals bearing tumors induced by Ha-SV, a protein with a molecular weight of 21,000 is present in cells transformed by Ha-SV (T. Y. Shih, M. O. Weeks, H. A. Young, and E. M. Scolnick, Virology, in press). A protein of the same size has also been found in cells transformed by Ki-SV (Shih et al., Virology, in press). This Ki-SV p21 protein cross-reacts immunologically with antisera from Ha-SV-induced tumors. Polypeptides of the same molecular weight and antigenic properties are found as in vitro translational products in systems programmed with the two viral RNAs (Shih et al., Virology, in press). (It may be noted that the p21 protein resulting from in vitro translation of Ki-SV RNA was not recognized in an earlier study [18]; the p50 Ki-SV translation product observed in that study is not antigenically related to the p21 proteins [Shih et al., Virology, in press].) The efficiency of in vitro translation for the p22 proteins appears to be greater for full-length viral RNAs than for shorter ones, suggesting that the proteins are coded near the 5′ end of the RNA molecules. The minimum coding sequence for a protein with a molecular weight of 21,000 is about 0.6 kb. The 5′ end sequence homology region between Ki-SV and Ha-SV is long enough to code for such proteins and could be the coding region. Nevertheless, these experiments do not rule out the possibility that the antigenically related p21's of Ki-SV and Ha-SV are coded by sequences other than the homology region at the 5′ end and possibly within the 4.36-kb region of sequence homology near the 3′ end.

(iv) As discussed in detail above, the interpretation of the rat endogenous 30S RNA/Ki-SV heteroduplexes is complicated because of the uncertainty as to whether the rat endogenous RNA is homogeneous in sequence. Nevertheless, the overall structural features indicate that there are mixed regions of sequence homology and nonhomology at both the 5′ and the 3′ ends. Furthermore, there is a large (4-kb) region of homology between the Ki-SV and the rat 30S RNA in the center of the genomes, with only the very small nonhomology feature B. Because the rat 30S-related genes are present as multiple copies in the rat genome, the regions of nonhomology are of interest since they may reflect the fact that each virus was formed by recombination with related but distinct copies of the 30S-related genes (19).

The present results show that the 3.36-kb sequence homology region between Ki-SV and Ha-SV, starting 1.1 kb from the 3′ end of the genome, is present in the rat endogenous 30S RNA, with the exception of the 0.2-kb feature B. We cannot determine unambiguously whether the 0.7-kb sequence homology region between Ki-SV and Ha-SV at the very 5′ end of the genomes is also present in the rat endogenous 30S genome. Therefore, further study is needed to identify the regions of the Ha-SV and Ki-SV genomes which code for the p21 proteins and to determine whether these sequences are present in the rat endogenous 30S RNA.

ACKNOWLEDGMENTS

This research was supported by Public Health Service contract NO1 CP 43306 from the Division of Cancer Cause and Prevention, National Cancer Institute, to N.D., by Public Health Service grants CA 16561 to I.M.V., CA 21408 to I.M.V., and CA 16113 to M.L. from the National Cancer Institute, and by American Cancer Society research grant VC-176 to M.L.

LITERATURE CITED

1. **Aaronson, S. A., and C. Weaver.** 1971. Characterization of murine sarcoma virus (Kirsten) transformation of mouse and human cells. J. Gen. Virol. **13**:245–252.
2. **Anderson, G. R., and K. C. Robbins.** 1976. Rat sequences of the Kirsten and Harvey murine sarcoma virus genomes: nature, origin, and expression in rat tumor RNA. J. Virol. **17**:335–351.
3. **Baily, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**:75–85.
4. **Bender, W., N. Davidson, K. L. Kindle, W. C. Taylor, M. Silverman, and R. Firtel.** 1978. The structure of M6, a recombinant plasmid containing Dictyostelium DNA homologous to actin messenger RNA. Cell **15**: 779–788.
5. **Benveniste, R. E., and E. M. Scolnick.** 1973. RNA in mammalian sarcoma virus transformed nonproducer cells homologous to murine leukemia virus RNA. Virology **51**:370–382.

6. **Chien, Y-h., I. M. Verma, T. Y. Shih, E. M. Scolnick, and N. Davidson.** 1978. Heteroduplex analysis of the sequence relations between the RNAs of mink cell focus-inducing and murine leukemia viruses. J. Virol. **28**:352–360.
7. **Donoghue, D. J., E. Rothenberg, N. Hopkins, D. Baltimore, and P. A. Sharp.** 1978. Heteroduplex analysis of the nonhomology region between Moloney MuLV and the dual host range derivative HIX virus. Cell **14**: 959–970.
8. **Harvey, J. J.** 1964. An unidentified virus which causes the rapid production of tumors in mice. Nature (London) **204**:1104–1105.
9. **Kirsten, W. H., and L. A. Mayer.** 1967. Morphologic responses to a murine erythroblastosis virus. J. Natl. Cancer Inst. **39**:311–335.
10. **Lai, M. M., and S. Hu.** 1978. In vitro synthesis and characterization of full and half genome length complementary DNA from avian oncoviruses. Nature (London) **271**:481–483.
11. **Maisel, J., M. Klement, M. C. Lai, W. Ostertag, and P. H. Duesberg.** 1973. RNA components of murine sarcoma and leukemia viruses. Proc. Natl. Acad. Sci. U.S.A. **70**:3536–3540.
12. **Rothenberg, E., D. J. Donoghue, and D. Baltimore.** 1978. Analysis of a 5′ leader sequence on murine leukemia virus 21S RNA. Heteroduplex mapping with long reverse transcriptase products. Cell **13**:435–451.
13. **Roy-Burman, P., and V. Klement.** 1975. Derivation of mouse sarcoma virus (Kirsten) by acquisition of genes from the heterologous host. J. Gen. Virol. **28**:193–198.
14. **Scolnick, E. M., R. J. Goldberg, and D. Williams.** 1976. Characterization of rat genetic sequences of Kirsten sarcoma virus: distinct class of endogenous rat type C viral sequence. J. Virol. **18**:559–566.
15. **Scolnick, E. M., and W. P. Parks.** 1974. Harvey sarcoma virus: a second murine type C sarcoma virus with rat genetic information. J. Virol. **13**:1211–1219.
16. **Scolnick, E. M., E. Rands, D. Williams, and W. P. Parks.** 1973. Studies on the nucleic acid sequences of Kirsten sarcoma virus: a model for formation of a mammalian RNA-containing sarcoma virus. J. Virol. **12**:458–463.
17. **Scolnick, E. M., D. Williams, J. Maryak, W. Vass, R. J. Goldberg, and W. P. Parks.** 1976. Type C particle-positive and type C particle-negative rat cell lines: characterization of the coding capacity of endogenous sarcoma virus-specific RNA. J. Virol. **20**:570–582.
18. **Shih, T. Y., D. R. Williams, M. O. Weeks, J. M. Maryak, W. C. Vass, and E. M. Scolnick.** 1978. Comparison of the genomic organization of Kirsten and Harvey sarcoma viruses. J. Virol. **27**:45–55.
19. **Shih, T. Y., H. A. Young, J. M. Coffing, and E. M. Scolnick.** 1978. Physical map of the Kirsten sarcoma virus genome as determined by fingerprinting RNase T1-resistant oligonucleotides. J. Virol. **25**:238–252.
20. **Tsuchida, N., R. V. Gilden, and M. Hatanaka.** 1974. Sarcoma-virus related RNA sequences in normal rat cells. Proc. Natl. Acad. Sci. U.S.A. **71**:4503–4507.